



Методи чисельної оптимізації

Лекція 3: Теоретичні основи оптимізації у глибокому навчанні

Кочура Юрій Петрович
iuriy.kochura@gmail.com
[@y_kochura](#)

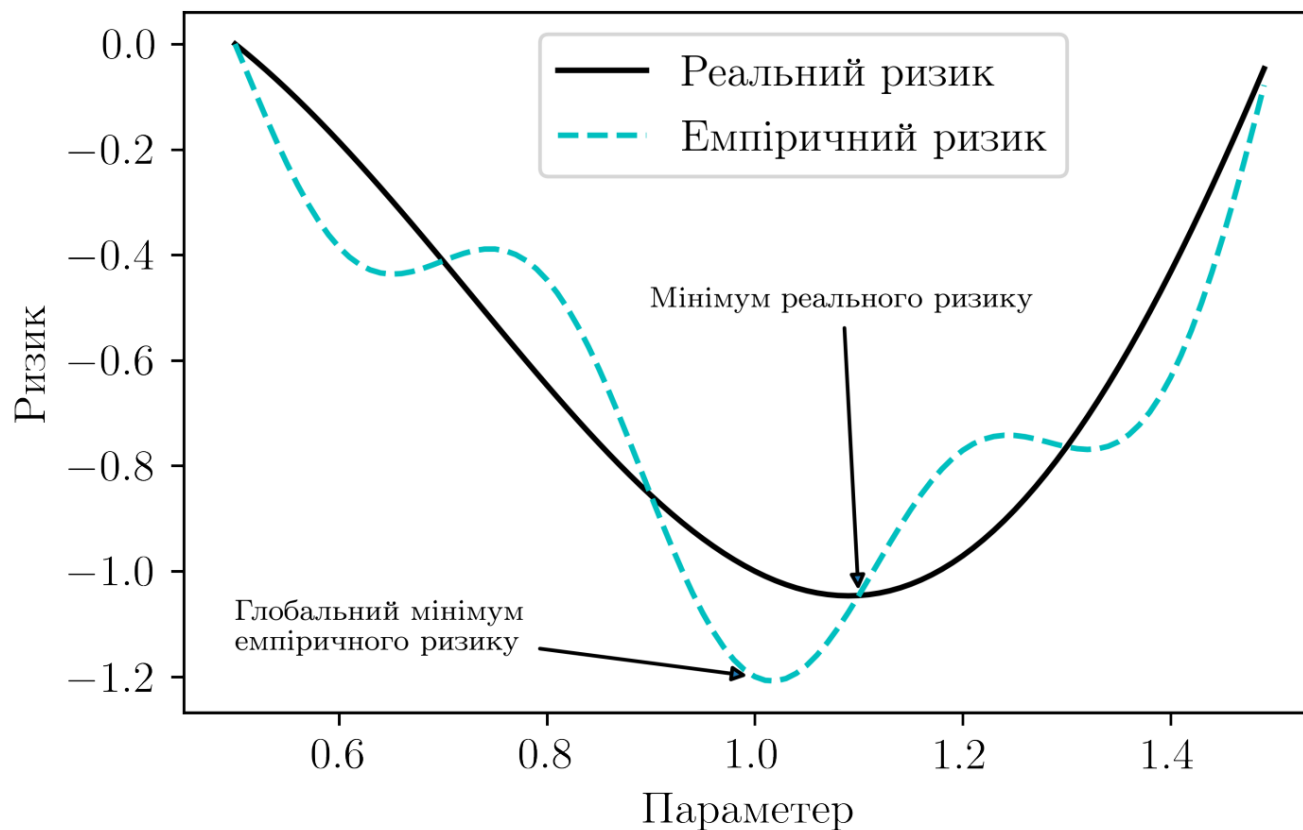
Сьогодні

- 🎤 Мета оптимізації
- 🎤 Проблеми оптимізації у глибокому навчанні
- 🎤 Опуклість множини та цільової функції
- 🎤 Нерівність Єнсена
- 🎤 Властивості опуклих функцій

Мета оптимізації

Оптимізація vs Глибоке навчання

Оптимізація	Глибоке навчання
Мінімізація цільової функції	Побудова адекватної моделі
Орієнтація на зменшення помилки на навчальних даних	Орієнтація на зменшення помилки узагальнення (на нових даних)
Використовує втрату на навчальних даних	Працює з обмеженими даними та прагне до узагальнення
Фокус: алгоритм	Фокус: модель + здатність узагальнювати
Ризик: локальні мінімуми	Ризик: перенавчання



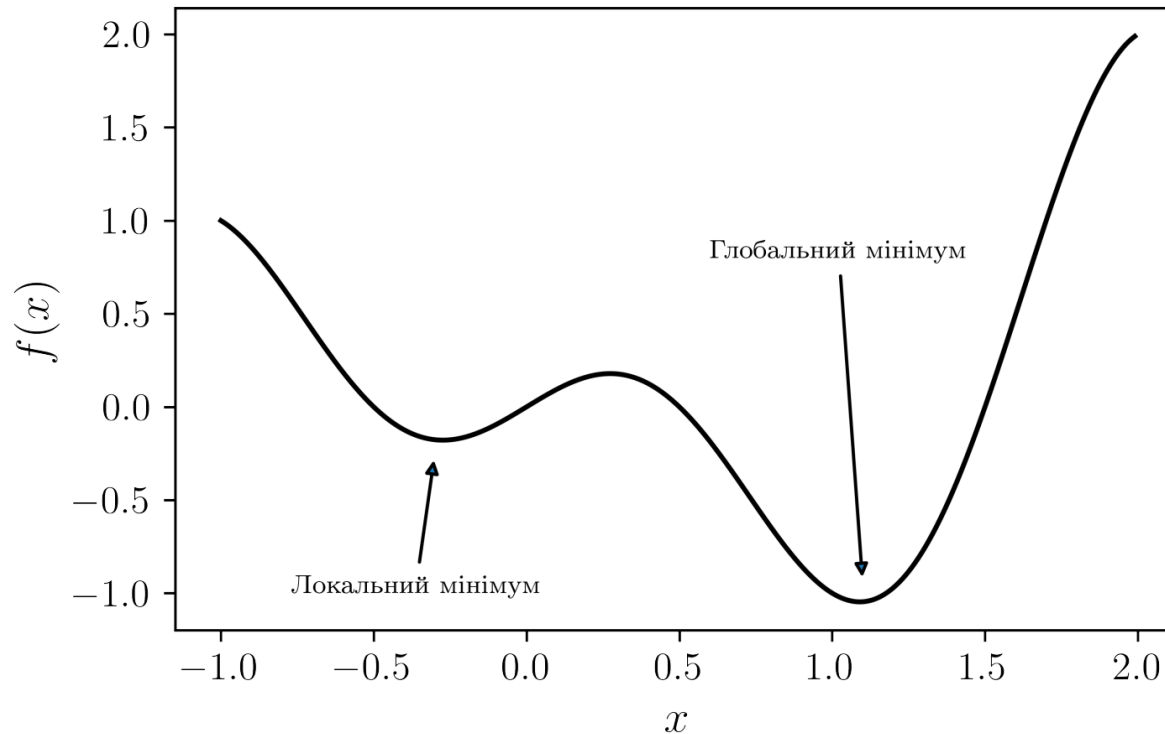
Мінімум емпіричного ризику на навчальному наборі даних може відрізнятися від мінімуму реального ризику (помилки узагальнення). Це ілюструє розбіжність між оптимізацією на навчальних даних та здатністю моделі узагальнювати на нових даних.

Проблеми оптимізації у глибокому навчанні

- У глибокому навчанні більшість цільових функцій складні і не мають аналітичного розв'язку, тому використовуються чисельні алгоритми оптимізації.

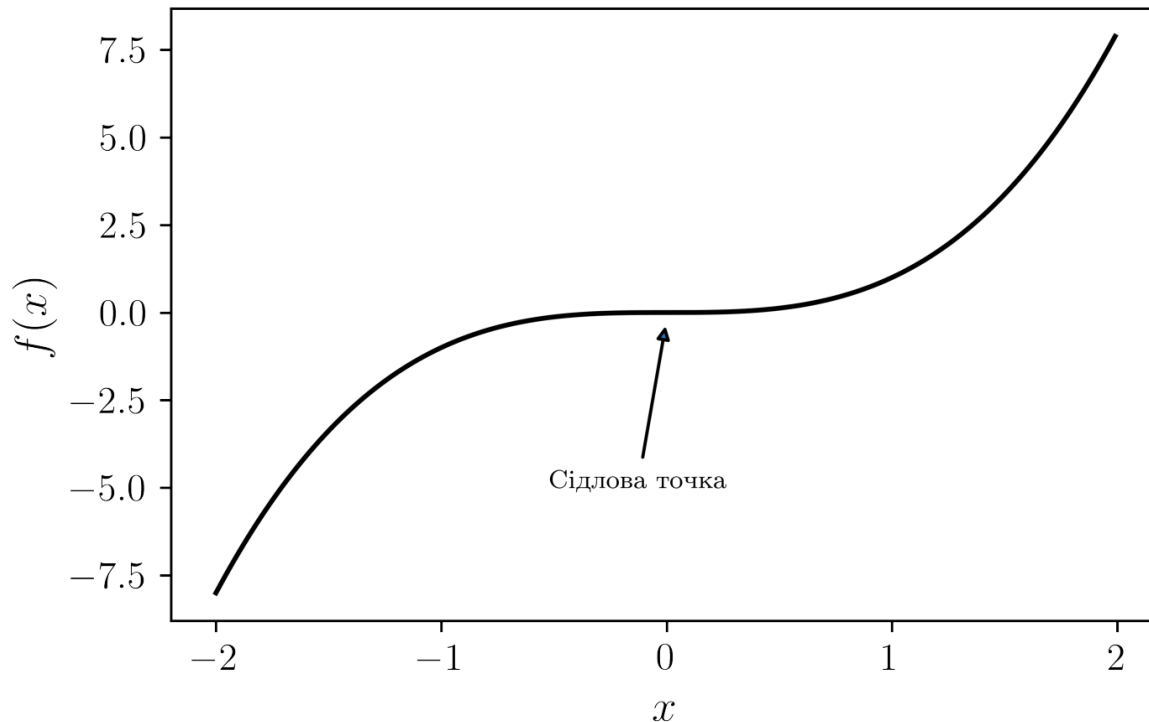
- У глибокому навчанні більшість цільових функцій складні і не мають аналітичного розв'язку, тому використовуються чисельні алгоритми оптимізації.
- Основні виклики: локальні мінімуми, сідлові точки та зникання градієнтів.

Локальний мінімум



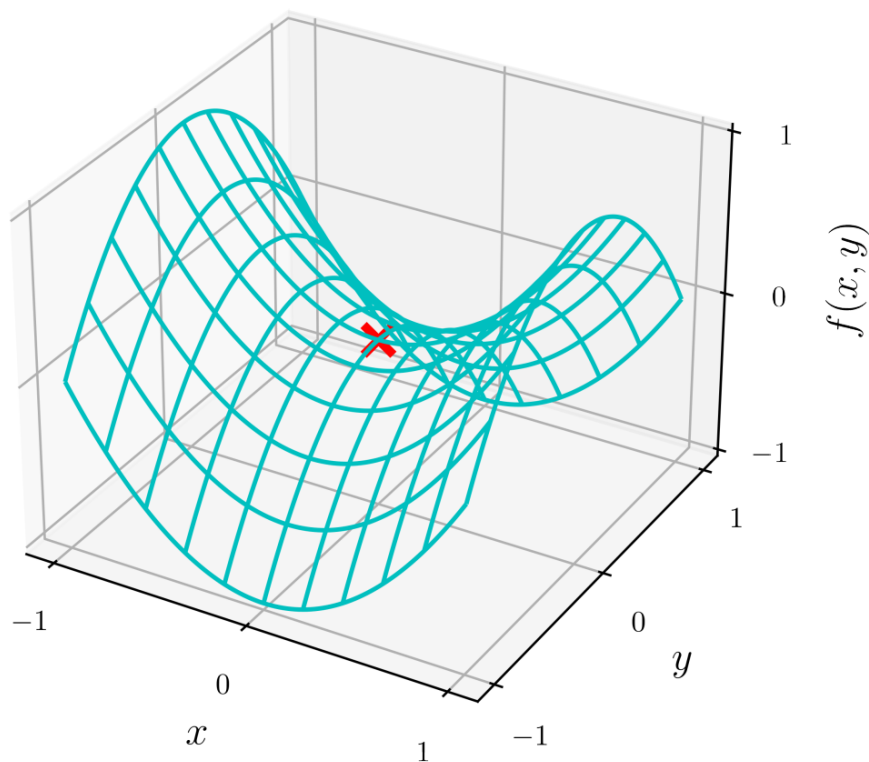
- Якщо значення $f(x^*)$ менше, ніж у всіх сусідніх точках, то x^* — локальний мінімум.
- Якщо значення $f(x^*)$ мінімальне на всій області визначення, то x^* — глобальний мінімум.

Сідлові точки (точки перегину)



Сідлова точка — це будь-яке місце, де усі градієнти функції зникають, але яке не є ні глобальним, ні локальним мінімумом.

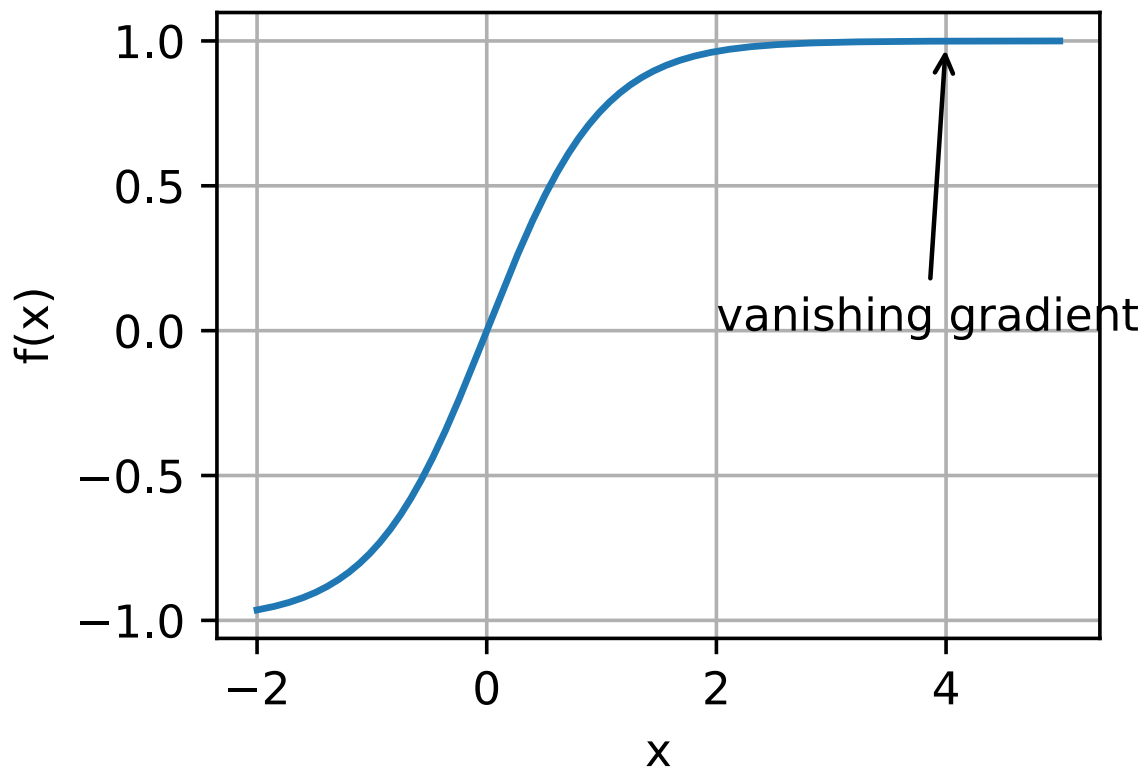
Сідлова точка **2D**-функції



$$f(x, y) = x^2 - y^2$$

Сідлова точка в $(0, 0)$.

Зникання градієнта



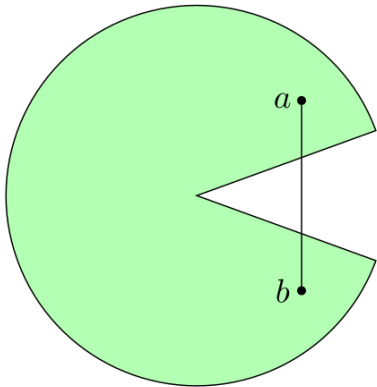
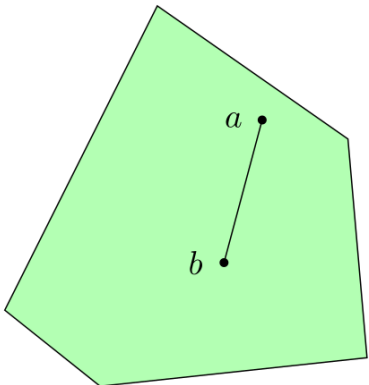
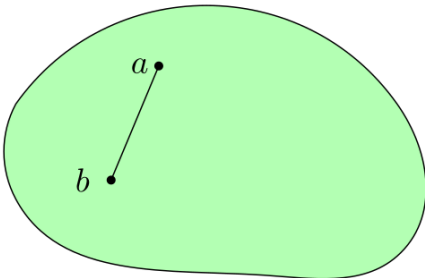
Опуклість множини та цільової функції

Опуклість множини

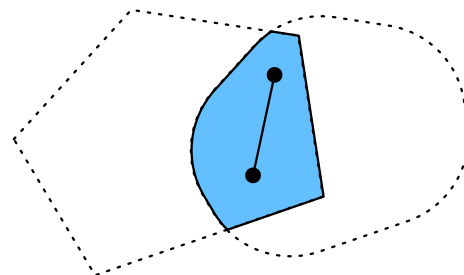
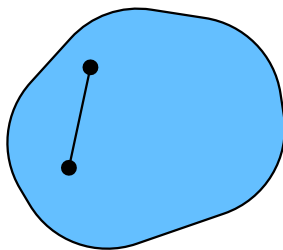
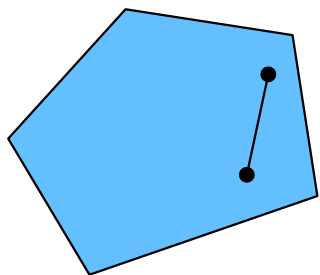
Поняття опуклої множини лежить в основі визначення опуклої функції.

Множина \mathcal{X} у векторному просторі є опуклою, якщо для будь-яких $a, b \in \mathcal{X}$, відрізок, що з'єднує a та b також знаходиться в \mathcal{X} . Математично це означає, що для усіх $\lambda \in [0, 1]$ маємо:

$$\lambda a + (1 - \lambda)b \in \mathcal{X}$$

Неопукла множина	Опукла множина	Опукла множина
		

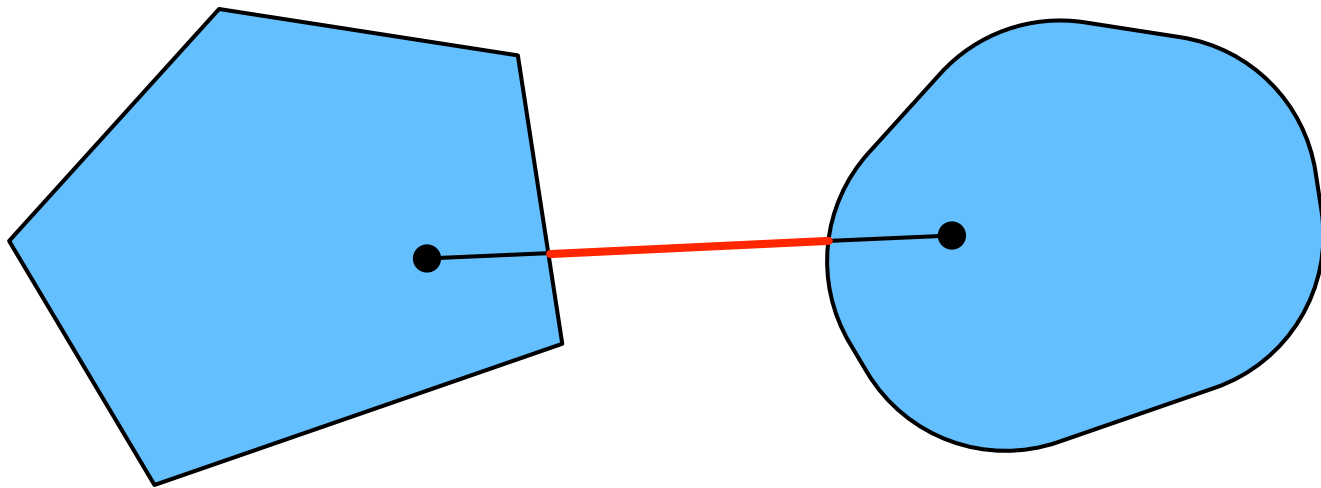
Опуклість множини



Якщо A і B — опуклі множини, то:

$A \cap B$ — теж опукла множина.

Опуклість множини



Якщо A і B — опуклі множини, то:

$A \cup B$ — може бути не опуклим.

Опуклість цільової функції

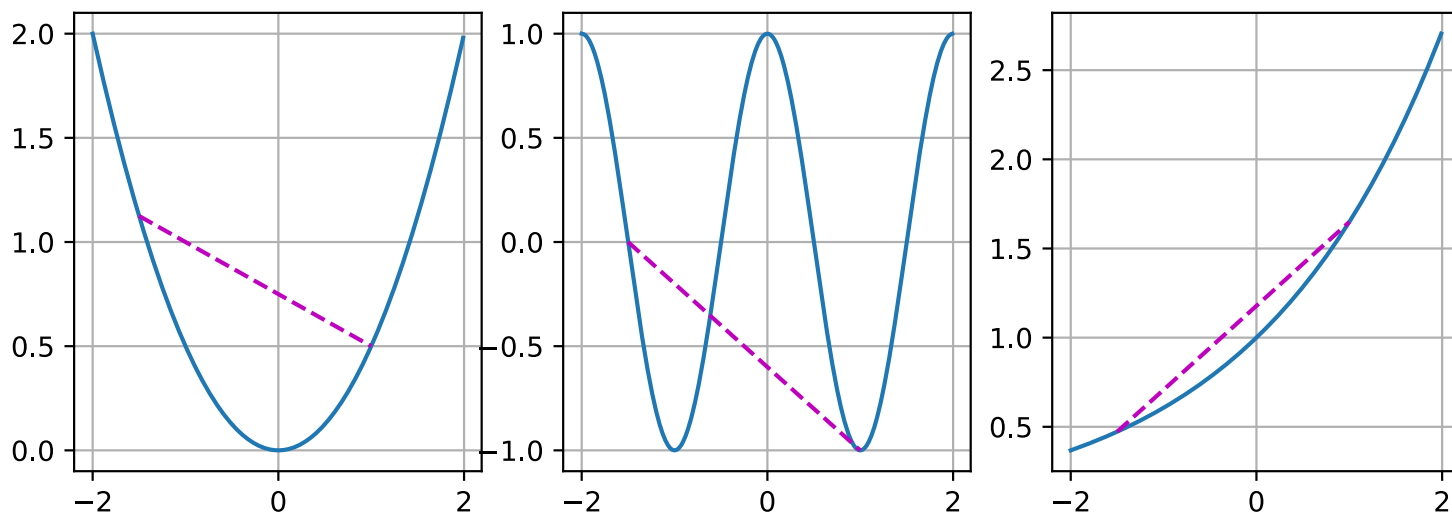
Визначення опуклої функції можна дати на основі поняття опуклої множини. Нехай дано опуклу множину \mathcal{X} , тоді функція $f : \mathcal{X} \rightarrow \mathbb{R}$ є опуклою, якщо для будь-яких $a, b \in \mathcal{X}$ та для усіх $\lambda \in [0, 1]$ виконується:

$$\lambda f(a) + (1 - \lambda)f(b) \geq f(\lambda a + (1 - \lambda)b).$$

Геометричне пояснення:

- Візьмемо дві точки на графіку функції: $(a, f(a))$ та $(b, f(b))$.
- Проведемо пряму між цими точками (хорду).
- Для опуклої функції будь-яка точка на хорді лежить над або на графіку функції.

Опуклість цільової функції



Нерівність Єнсена

Нерівність Єнсена

Для опуклої функції f та випадкової величини X із невід'ємними вагами α_i , де $\sum_i \alpha_i = 1$:

$$\sum_i \alpha_i f(x_i) \geq f\left(\sum_i \alpha_i x_i\right)$$

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$$

- Використовується для оцінки складних виразів через простіші.
- Дозволяє отримати нижню межу або наближення, яке легше аналізувати та оптимізувати.

Властивості опуклих функцій

Властивості опуклих функцій

- Локальний мінімум — глобальний мінімум
 - Мінімізуючи функцію ми не «застрягаємо» у локальному мінімумі.
 - Може бути багато глобальних мінімумів або їх відсутність (функція може асимптотично наближатися до мінімуму, але не досягати його).

Властивості опуклих функцій

- Локальний мінімум — глобальний мінімум
 - Мінімізуючи функцію ми не «застрягаємо» у локальному мінімумі.
 - Може бути багато глобальних мінімумів або їх відсутність (функція може асимптотично наближатися до мінімуму, але не досягати його).
- Підмножина опуклих функцій — також є опуклою.

Властивості опуклих функцій

- Локальний мінімум — глобальний мінімум
 - Мінімізуючи функцію ми не «застрягаємо» у локальному мінімумі.
 - Може бути багато глобальних мінімумів або їх відсутність (функція може асимптотично наближатися до мінімуму, але не досягати його).
- Підмножина опуклих функцій — також є опуклою.
- Опуклість і другі похідні:

$$f \text{ опукла} \iff f''(x) \geq 0 \quad \forall x$$

